

EXTRACT DATA FROM PDF TO A DATABASE

You have business documents you get in pdf format: invoices, work orders, purchase orders and others. Sometimes data is in the pdf as a table or documents were scanned into a pdf. They hold data you need to process in your ERP or other database-driven information system.

How to convert these PDF documents into usable data in your database?

You use Docparser, that's how!

Docparser is a leading PDF converter with some processing muscle and a few friends to get the heavy-lifting of data intake done for you.

YOUR CHOICES FOR CONVERTING THE PDF DATA

This post refers to MySQL database, where Docparser is the first step to building your PDF to MySQL database converter.

Keep in mind that Docparser has no requirements on database vendors or scripting languages. You are free to use the database and language of your choice.

1. [Set up parsing rules](#) and [import your files](#) for each type of document you want to bring in. This step is required no matter where data goes after capture. You will need Docparser to get the data out of the PDF and ready for your database.
2. Determine which method you will use to move PDF data to the MySQL database.
 - a. Google Sheet or csv file manual download then import to the MySQL database.
 - b. Partner Integration to move the data from Docparser to your database. This can be through Zapier, Stamplay, or Workato, each allowing you to create the workflows you need.
 - c. Build a custom script via the API to move data from Docparser to your database.

OPTIONS FOR CONVERTING PDF DATA TO THE MYSQL DATABASE

Each different type of document you process requires its own parsing rule. If you have 2 vendors using the same template for invoices, you can use the same parser for both. Clients often use a separate parser for each vendor for clarity.

THE FIRST OPTION REQUIRES THE MOST MANUAL INTERVENTION, BUT IS QUICK TO IMPLEMENT.

Use Docparser functionality to convert the PDF data to a csv file or Google Sheet ready for upload into the database.

You build a parser for each different type of document you process and assign a specific method for handling the data.

Docparser can create csv files direct from PDF documents and has a built-in method for [exporting data to a Google Sheet](#). Either one can be the default processing method for your parser. The data moves as soon as file processing completes.

MySQL has an [upload function](#) and will require you to download the file and then import to the database.

While this does require manual intervention, it is a good way to move information to your MySQL database while you are building all the parsers for your different types of documents.

IN THE NEXT CHOICE WE ADD A WEBHOOKS INTEGRATION TO OUR WORKFLOW TO MOVE THE PDF DATA INTO OUR MYSQL DATABASE.

As soon as Docparser processes the incoming file, data posts to the integration platform you have identified for that parser. The information loads to your MySQL database through the Integration Partner.

[Here's information on the Zapier and Docparser workflow.](#)

[Workato has several pre-built recipes available for Docparser.](#)

Stamplay uses a [Webhook Integration](#) to get the Docparser data.

IN THE THIRD METHOD, USE THE DOCPARSER API TO EXTRACT PDF DATA TO A MYSQL DATABASE.

A developer can build a custom script to pull the parsed data and move it into the correct database location. This API has your own signature and sign on, tied to your data and files.

The API offers a great deal of functionality including:

- Identify all defined parsers, by ID and name.
- Upload documents for parsing, via HTML form or an accessible URL
- Apply a unique identifier to any document which you submit

- Receive data via a Webhook Integration to your application, a permanent download link, or by polling and fetching from your API
- Fetch data for your API in single or multiple data sets

System and network efficiencies come with using the [Advanced Webhook Integration](#) to push parsed data to your MySQL database. Once a new document is parsed, it then sets off a trigger, eliminating polling activity and providing data to the database. This is usually complete within 1 to 3 minutes of document submission. From there, your MySQL database table populates immediately, by a timer or based on data volume levels.

WHERE TO START?

Some clients start with one method and build their next iteration to a different method. This is a good way to expedite your data capture while leveraging available tools and testing your process.

These are the different ways to convert a pdf to data in a MySQL database and Docparser can help simplify this process. The Docparser team is always here to help you get up and running as quickly as possible. Quit re-entering your data! Isn't it time you got some automation into your data intake?

Sign up for a free 30-day trial now and see how much easier your workday can be with Docparser.